Analysing performance of first year engineering students

Conference or Workshop Item

Analysing performace of first year engineering students

Zdenek Zdrahal
KMi, The Open University
Milton Keynes, UK
CIIRC, CTU in Prague
Prague, CZ
zdenek.zdrahal@open.ac.uk

Martin Hlosta
KMi, The Open University
Milton Keynes, UK
martin.hlosta@open.ac.uk

Jakub Kuzilek
KMi, The Open University
Milton Keynes, UK
CIIRC, CTU in Prague
Prague, CZ
jakub.kuzilek@open.ac.uk

ABSTRACT

Many students in the engineering disciplines do not complete their higher education degree and drop out. This problem is serious, especially for first-year university students. In this paper, we analyse how students earn the credits required for their successful completion of the first study year. Using the example of a European technical university with traditional classroom-based education, we identify three groups of students: those who pass, those who earn only enough credits for staying in the program, and those who fail. Important patterns can be found at the end of the first semester. We present a simple algorithm that identifies students who may benefit from early additional support, which would increase their chances of progression to the second year and improve the retention improvement for the university. The results are evaluated in four consecutive academic years. The data from years 2013/14 and 2014/15 have been used to develop and verify the prediction model. In study years 2015/16 and 2016/17 the model has been applied to predict at-risk students, where the university tutors intervened and provided additional support and a significant improvement was achieved.

Keywords

Student drop-out, learning analytics, intervention, progression, engineering education, STEM subjects, prediction of study results, ECTS credits, early exam period, first-year bachelor's program, traditional classroom-based university.

1. INTRODUCTION

According to Quinn, [3] in some EU countries between 20% and 54% of students fail to complete their degrees. In distance education, the percentage of students who fail to complete the degree is about 78%, see [4].

This paper analyses patterns of behaviour exhibited by the cohorts of first-year students. The aim of the analysis is to identify students at risk of failing as early as possible so that they can receive suitable support [1]. An anonymised dataset has been taken from the Faculty of Mechanical Engineering, at the Czech Technical University, which offers a three-year bachelor program, followed by a two-year master's

---

This KMi report is based on the LAK 2016 workshop paper [6] with the same authors and title. The content of the LAK 2016 paper has been updated and the results of interventions in 2015/16 and 2016/17 included.

program in Mechanical Engineering. The education is organised in the traditional classroom-based manner with lectures, tutorials and exams.

Data of 994 students registered in 2013/14 were used for developing the predictive model. This model was verified on 917 students registered in 2014/15. Based on the predictions, tutors intervened in 2015/16 and 2016/17.

2. STUDY ORGANISATION

2.1 Academic year

The academic year is divided into winter and summer semesters each 13 weeks long, and each followed by 6 week examination periods. At the end of the winter semester, before the start of the winter exam period, there is 1 week Christmas break. Prior to the winter and summer exam periods, still within the running semester, there are "early exam periods" of at least 4 weeks. The summer exam period is extended by 2 more weeks after the summer holiday. Students can earn ECTS credits in 20 courses, 12 of which belong to the Science, Technology, Engineering and Mathematics (STEM) group. These are the most important courses for acquiring the qualification [2] and therefore are rewarded by more credits. In the winter semester, students can earn up to 39 ECTS credits of which 32 are from STEM subjects, in the summer semester it is 36 credits in total of which 32 are STEM credits. For each course, the examiner offers multiple dates with a number of students who can register for the date and students can decide themselves how to schedule their exams.

2.2 Progression rules

Based on their performance, the student can achieve 3 different results:

- **Pass** – if the student earns at least 30 credits in the winter semester and 60 credits in total during the academic year. Such students successfully complete the academic year. Students who pass have to earn all credits from all STEM subjects.
- **Continuing** – if the student earns more than 15 credits in the winter semester and between 30 and 59 credits in total during the academic year. Such students can in the future take only the failed or missing courses and earn the corresponding credits.
- **Fail** – if the student earns fewer than 15 credits in the winter semester or fewer than 30 credits in total during the academic year. Such students are deregistered from the program.

In the academic year 2013/14, 994 students registered for first year courses and 330 of them failed. In 2014/15, 917 students registered and 382 failed.

This progression rate is obviously unsatisfactory for students, the university and society. Students waste time that is supposed to be used for improving their qualifications, the university loses financial resources and society suffers from the lack of qualified people needed in the industry.

For our analysis we define three student groups: **pass**, **continuing** and **fail** based on their results. In order to propose a possible improvement, we provide a deeper insight into the behaviour of these groups.

3. ANALYSING PATTERNS OF CREDIT EARNING

First of all, we have to realize that the ultimate assignment to the pass, continuing or fail group is done at the end of the academic year, with the exception of students who irreversibly fail in the winter semester by not earning the minimum of 15 credits. However, many students fluctuate between the fail, continuing and pass groups. The dynamics of these groups classified at the end of the academic year is shown in Figure 1.

On the horizontal axis are dates in the study year with important milestones highlighted by a vertical line. The vertical axis shows the average number of credits for each class. The total number of ECTS credits for different groups is shown in full lines, the STEM credits are depicted by dashed lines. Periods of the academic year are denoted as follows: (a) winter early exam period as a part of the winter term, (b) Christmas break, (c) winter exam period, (d)+(e) summer term, (e) summer early exam period, (f) summer exam period, (g) summer holiday, (h) end of the academic year. The beginning of the winter term is not shown as it is not important for our analysis.

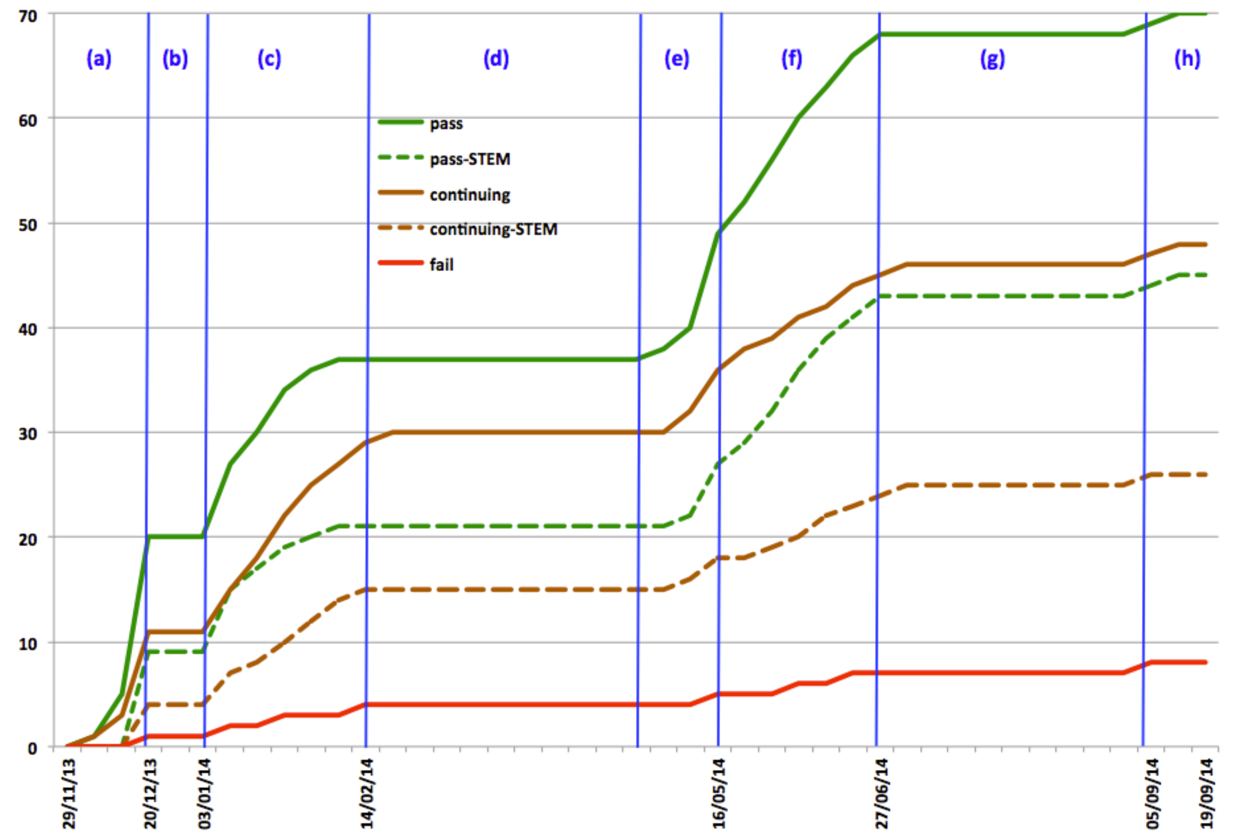

Figure 1. Average number of credits earned by different groups in time

3.1 Predicting the final result from the result of the winter semester

When the winter exam period is over and the winter semester results are known, it is possible to provide a crude estimate of the overall progression. The students who have not earned 15 credits in the winter semester have definitely failed and are deregistered. In the presented case there 446 such students. There are many reasons why these students fail. The rest of the cohort continues continue studying to finish at the end of the academic year in one of the groups: pass, continuing or fail.

The probability of a different final result depending on the number of credits earned at the end of the winter exam period is shown in Figure 2. The bars indicate the percentage of students who at the and of the study year pass, continue or fail given the number of credits earned at the end of the winter semester. For example, out of the students who earn 24 credits at the end of the winter semester 10% will pass, almost 70% will be continuing and about 20% will fail. According to the rules, all students with fewer credits than 15 have failed. Some students who failed could have been saved if an early enough action had been taken. It is obvious in Figure 1 that the differentiation into pass, continuing and fail groups exists already at the start of the winter exam period. The credits earned at the start of the winter exam period are awarded both for early exams but also for activities, assignments, and test carried out during the winter semester.

The pattern found in the data is that all students without any credit at the start of the winter exam period fail. There are no exceptions to this rule - it applies although there are plenty of opportunities to earn the sufficient number of credits during the winter exam period.

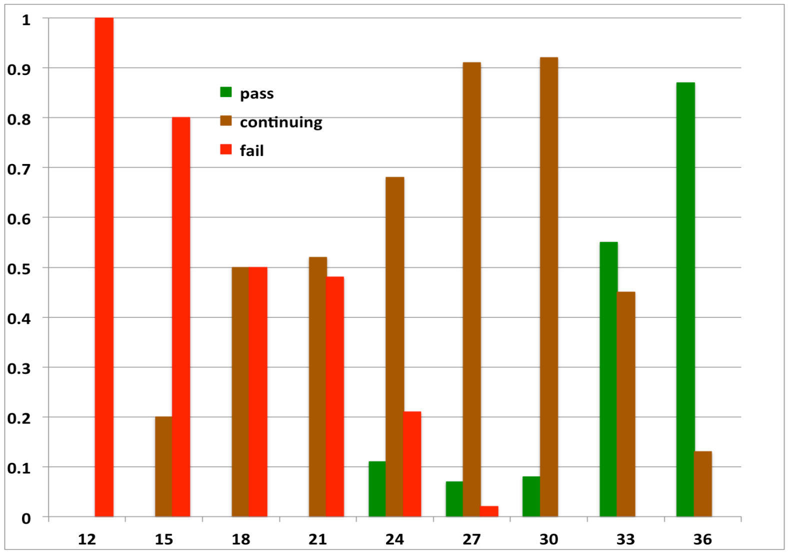

Figure 2. Probabilities of the final result based on the number of ECTS credits earned in the winter semester.

If it is possible to select at-risk students before the start of the regular exam period, it might be possible to provide them with additional support. One of the important challenges the students face are how to organize their first difficult, esp. STEM exams to minimize the number of failed opportunities. The university is prepared to assign a tutor to each small student groups to support them through this period, providing that there is a chance that the effort is efficient and meaningful.

3.2 Selecting students for interventions

The algorithm for selecting students who may need additional support is based on the assumption that the pattern of student behaviour persists across different school years; therefore the results calculated from historical data are applicable in the future.

Our goal is to select the group of at-risk students based on the knowledge of the number of credits the students earned at the start of the winter exam period (or any other fixed date) and the final classification of students into the above mentioned three groups. Let *N(x)* be the number of students who earned x credits, and N the total number of students. Similarly, we denote $N(pass\,\&\,x)$ the number of students who have earned (at the start of the winter exam period) exactly x credits and at the end of the academic year belong to the *pass* group. We can define probability $P(pass\,\&\,x) = \frac{N(pass\,\&\,x)}{N}$ and the conditional probability $P(pass|x) = \frac{P(pass\,\&\,x)}{P(x)}$. Similar probabilities can be defined for the *continuing* and *fail* groups. $P(pass|x) + P(continuing|x) + P(fail|x) = 1$ holds for any x. The probabilities for the three groups and the different number of credits at the start of the exam period are shown in Figure 3. The goal is to estimate the allocation of students to the groups based on early information about the credit distribution. Let us define $\mathrm{e(pass|x)} = \mathrm{P(continuing|x)} + \mathrm{P(fail|x)}$ the error caused by not assigning students with x credits to the *pass* group. Similarly, we can define the error function for the other two groups.

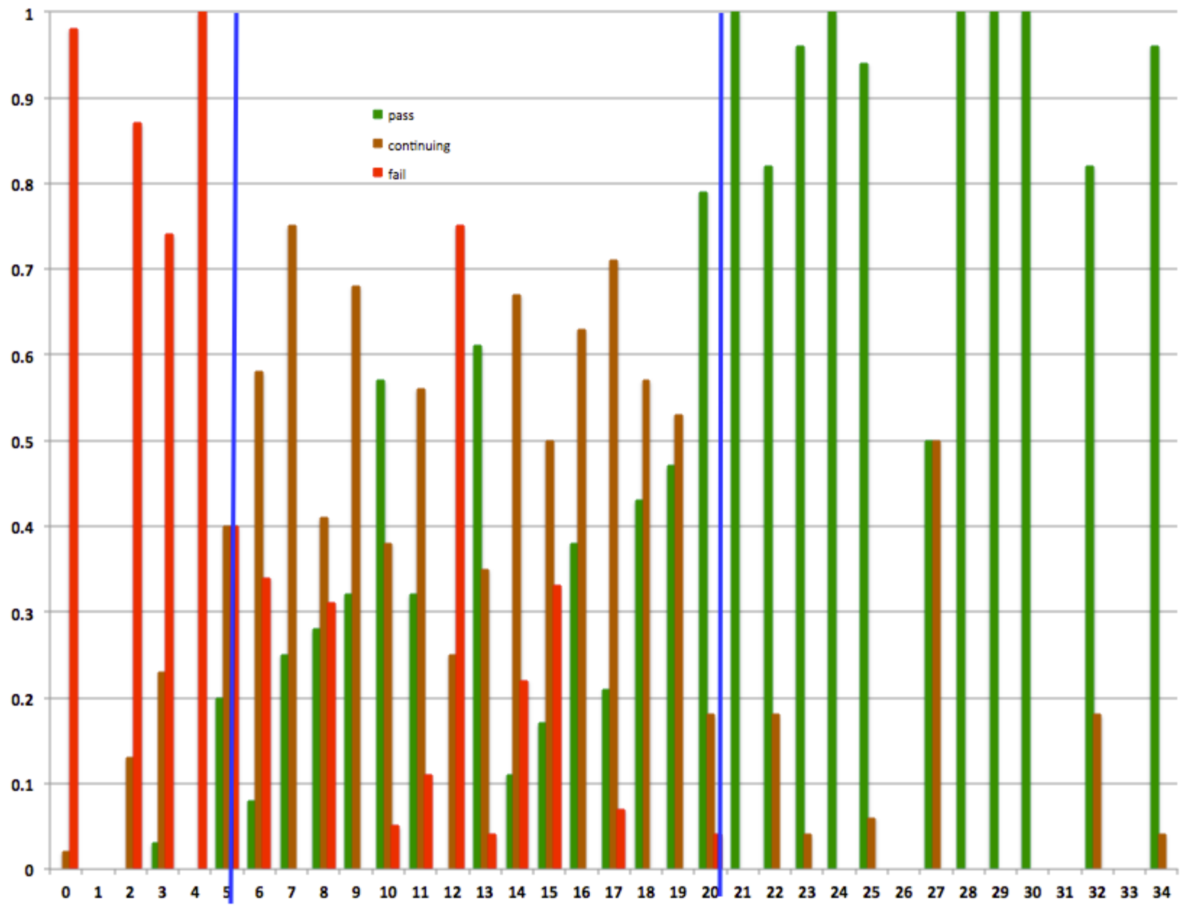

Figure 3. Probabilities of final results for a different number of credits earned at the start of winter exam period

It is reasonable to assume monotonicity in the assignment to the groups: up to some *x1* the students will be classified as *fail*, then from *x1* to *x2* as *continuing* and above *x2* as *pass*. The values *x1* and *x2*, depicted in Figure 3 as blue lines are selected to minimize the expression

$$E(x1, x2) = e_{fail}(x1) + e_{cont}(x1, x2) + e_{pass}(x2) ,$$

where

$$e_{fail}(x1) = \sum_{0}^{x1} e(fail|x) ,$$
$$e_{cont}(x1, x2) = \sum_{x1}^{x2} e(continuing|x), \text{ and}$$
$$e_{pass}(x2) = \sum_{x2}^{72} e(pass|x).$$

The value 72 in the last sum is the highest number of credits that can be earned. The terms measure misclassification of the *fail*, *continuing* and *pass* classes, respectively.

In Figure 3, some values of *x* are missing. There is no course awarded by 1 credit and therefore value *x*=1 is not in the histogram. The combinations 26, 31 and 33 credits could be achieved but do not exist in the analysed dataset.

The minimum of *E(x1, x2)* has been achieved for *x1* = 5 and *x2* = 20. The error $e_{cont}(x1, x2)$of the *continuing* class contributes by about 50% to the value of *E(x1, x2)*.

4. IMPLEMENTATION

The StudentAnalyse system is the model and dashboard for tracking and analysing student credits and has been implemented in the R language. The system has been developed and deployed in three phases. First, the predictive model has been developed from student data of 2013/14 and verified using 2014/15 data. In 2015 the model and the dashboard were implemented in the R language. Then, in 2015/16 and 2016/17 the model was used for improving retention.

A screenshot of the dashboard is shown in Figure 4. The use of the whole system was made available to the Faculty of Mechanical Engineering, CTU in 2015/16.

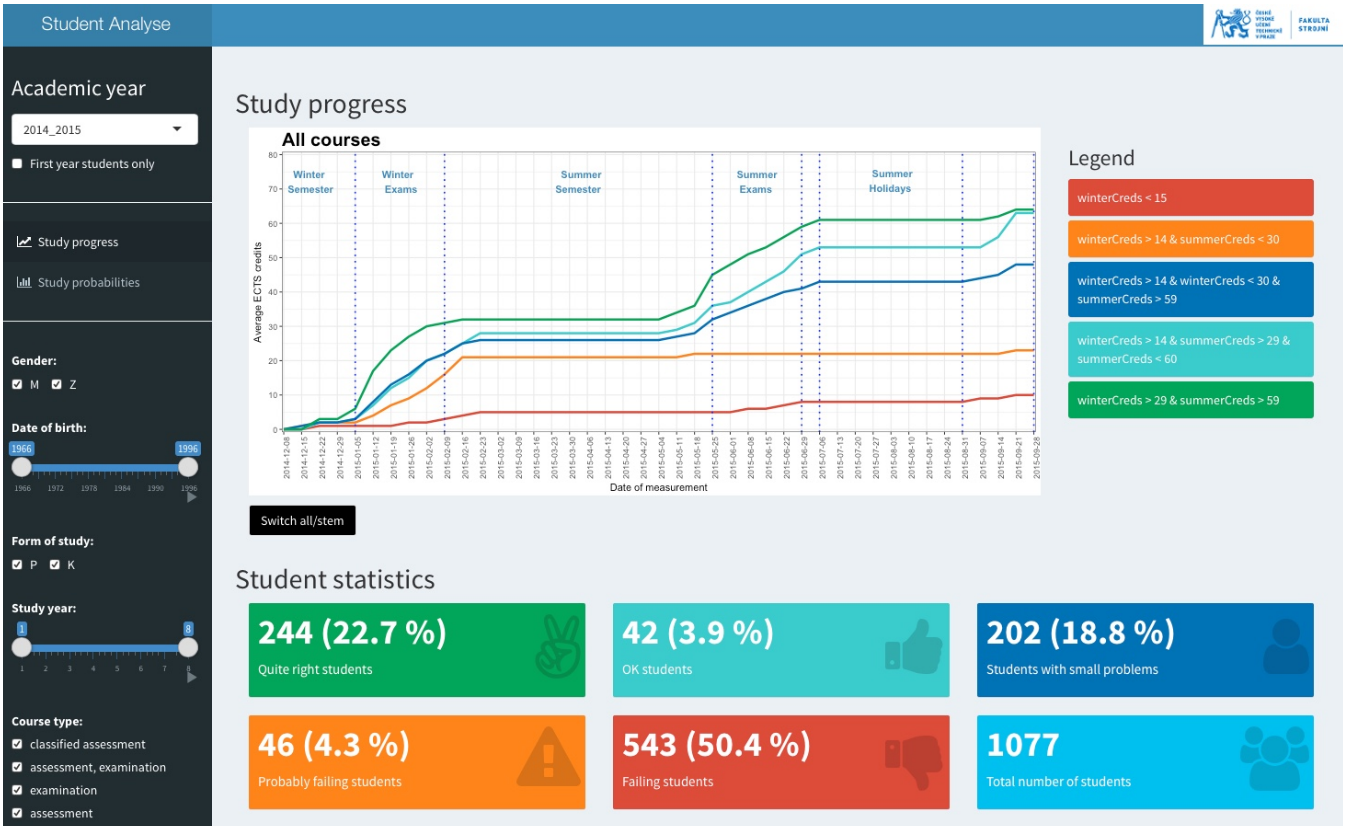

Figure 4. StudentAnalyse dashboard

5. SYSTEM EVALUATION

5.1 Predictions, interventions, and financial implications

The accuracy (i.e. precision) of the prediction was calculated as the ratio of the number of students predicted correctly to the number of all students predicted as being a member of the class. The results are as follows:

In class predicted as *fail (x<5)*, 100% students were predicted correctly. In the class predicted as *continuing (5 ≤ x ≤ 20)*, 50% students were at the end of the academic year in the class *continuing*, 31% were in *pass* and 19% students *failed*. In the class predicted as *pass* (x>20), 92% students were in the class *pass*, 8% were *continuing* and no student in reality failed. These results apply for study year 2013/14, for other years the values are similar.

The predictions of the *fail* and *pass* classes are accurate, the precision of the *continuing* class is significantly lower. This result is in accordance with our intuition when we look at Figure 3. The fact that the students belonging eventually to the *fail* or *pass* classes have been misclassified as *continuing* is not significant. These are borderline cases and they should be the target of interventions. The impact of predictions and interventions is evaluated both as the number of failed students and in terms of financial losses for the university.

The Czech government contributes to the university for teaching 57,750 CZK = £1,862 per student and year [5], the exchange rate used in the calculation is 1£ = 31CZK

The following results have been achieved:

- 2013/14...994 students registered, predictive model development, no intervention, 33.2% students failed, financial loss = 19,057,962 CZK

- 2014/15...917students registered, predictive model validation, no intervention, 41.1% students failed, financial loss = 21,765,224 CZK

- 2015/16 ...769 students registered, predictions calculated before the start of exam period, 3 interventions by tutors, 16.9%, i.e. 130 students failed, financial loss = 7,507,500 CZK

- 2016/17...688 students registered, early tutor interventions: 11Oct16, 8Dec16, 14 Dec 16 and 1 additional intervention 14 days before the end of the exam period, 7.99% failed, 8.72% withdrawn, i.e. in total 16.71%, i.e. 115 students have not completed, financial loss = 6,641,250 CZK

If we assume the same percentage

6. COMPARISON WITH THE PREVIOUS YEARS

The best results in the period prior the use of StudentAnalyse were achieved in 2013/14 when only (!) 33.2% students failed in the first study year. Assuming the same failure rate in 2015/16, 255 students would have failed and the financial loss for the university would have been 14,726,250 CZK. The net benefit for that year was therefore 7,221,003 CZK.

Similarly, in 2016/17, 228 students would have failed resulting in the financial losses of 13,167,000 CZK. The financial benefit is 6,527,783 CZK.

The total financial benefit for the university in the period 2015 – 2017 is therefore 13,748,786 CZK, i.e. about £443,500.

7. CONCLUSIIONS

The most important findings are that it is possible to predict the final result of the first year students at a traditional classroom-based university from their performance at the beginning of the study year, prior to the first exam period and that well-directed interventions do make a difference. Since in the context of the Open University, OUAnalyse predicts success in TMA submissions with a good precision and recall well before the cut-off date, it is likely that better organised and well targeted interventions would significantly improved student drop out with the obvious financial benefits for the university.

Both of the results achieved in 2015/16 and 2016/17 at the faculty of Mechanical Engineering, CTU are significantly higher than maximum 10% of retention shown in Table 1. We do not claim that the actual improvement of about 50% at CTU can be ascribed only to the well organised and systematically applied interventions, but we believe that this is a very important factor that also has the potential to improve the retention at the OU.

8. REFERENCES

[1] E. Aguiar, H. Lakkaraju, N. Bhanpuri, D. Miller,B. Yuhas, and K. L. Addison. Who, when, and why:A machine learning approach to prioritizing studentsat risk of not graduating high school on time. InProceedings of the Fifth International Conference onLearning Analytics And Knowledge, LAK '15, pages93{102, New York, NY, USA, 2015. ACM.

[2] Chen X. and Soldner M. 2014. *STEM Attrition: College Students' Paths Into and Out of STEM Fields Statistical Analysis Report* National Center for Education Statistics . NCES 2014-001. U.S. DEPARTMENT OF EDUCATION.

[3] Quinn, J. 2013. *Drop-out and Completion in Higher Education in Europe among students from under-represented groups*. European Commission.

[4] Simpson, O. 2010. *22% - can we do better? Supporting students at a distance*. From Articles on Distance Education: http://www.94669.mrsite.com/graduating high school on time. InProceedings of the Fifth International Conference onLearning Analytics And Knowledge, LAK '15, pages93{102, New York, NY, USA, 2010.

[5] MSMT (2017) Czech Ministry of Education, Youth and Sport, Section for Higher Education, Science and Research, Personal communication, 23 March 2017, information provided Mr Petr Valasek by phone.

[6] Zdrahal Z., Hlosta M. and Kuzilek J. (2016). Analysing performace of first year engineering students. Workshop paper. In: A. Wolff, Z. Zdrahal (eds.): Proceedings of the Data literacy for Learning Analytics Workshop, LAK 2016 Conference, Edinburgh, UK, 26-Apr-2016, published at http://ceur-ws.org